

OPTIMIZED EDGE APPEARANCE PROBABILITY FOR COOPERATIVE LOCALIZATION BASED ON TREE-REWEIGHTED NONPARAMETRIC BELIEF PROPAGATION

Vladimir Savic¹, Henk Wymeersch², Federico Penna³, Santiago Zazo¹

¹Universidad Politecnica de Madrid, Spain, Email: {vladimir, santiago}@gaps.ssr.upm.es

²Chalmers University of Technology, Gothenburg, Sweden, Email: henk.wymeersch@ieee.org

³Politecnico di Torino, Italy, Email: federico.penna@polito.it

ABSTRACT

Nonparametric belief propagation (NBP) is a well-known particle-based method for distributed inference in wireless networks. NBP has a large number of applications, including cooperative localization. However, in loopy networks NBP suffers from similar problems as standard BP, such as over-confident beliefs and possible non-convergence. Tree-reweighted NBP (TRW-NBP) can mitigate these problems, but does not easily lead to a distributed implementation due to the non-local nature of the required so-called edge appearance probabilities. In this paper, we propose a variation of TRW-NBP, suitable for cooperative localization in wireless networks. Our algorithm uses a fixed edge appearance probability for every edge, and can outperform standard NBP in dense wireless networks.

Index Terms— Cooperative localization, nonparametric belief propagation, tree-reweighted belief propagation, wireless networks

1. INTRODUCTION

Cooperative localization is an important problem in wireless networks, as the availability of positional information can enable many applications, such as search-and-rescue, asset tracking, and indoor navigation. State-of-the-art algorithms for cooperative localization rely essentially on *nonparametric belief propagation* (NBP) [1, 3, 5]. NBP is a variation on BP, which is a well-known, low-complexity inference method, with applications in many fields, including artificial intelligence, wireless communication, coding theory, computer vision, and cognitive networks. For continuous random variables, messages can often not be computed or represented exactly. NBP was proposed in [1] to address this problem through a particle-based representation. Both BP and NBP can be interpreted as message passing algorithms on a graphical model, and are inherently distributed, thus lending themselves well to distributed inference problems, such as cooperative localization.

For most applications, the main problem of NBP (and also BP) is that in graphs with cycles, there are no guarantees on the quality of the marginal beliefs, nor on the convergence of message passing. Solutions to this problem include generalized belief propagation (GBP) [2], NBP over spanning trees (NBP-ST) [3], and tree-reweighted belief propagation (TRW-BP) [4]. GBP is able to improve on standard BP, but at high computational cost, making it unsuitable for large-scale networks. NBP-ST performs NBP on two (or

more) spanning trees, outperforming NBP only in highly connected networks, with small performance gains. Finally, TRW-BP considers distributions over all spanning trees, through a concept known as *edge appearance probabilities*. TRW-BP has the potential to outperform NBP, but relies on the availability of valid edge appearance probabilities, which are themselves hard to obtain in a distributed manner. Hence, TRW-BP in its original form is not suitable for a distributed implementation.

In this paper, we propose a variation of tree-reweighted NBP for cooperative localization in wireless networks, where we (i) consider *uniform* edge appearance probabilities (characterized by the scalar variable ρ); (ii) allow *invalid* edge appearance probabilities (i.e., not corresponding to any distribution over spanning trees). We evaluate the proposed method as a function of ρ in terms of two performance metrics: the root-mean square error (RMSE) of the position estimate and the Kullback-Leibler divergence (KLD) of the marginal belief with respect to the true marginal posterior. We propose an empirical function for the optimum edge appearance probability that minimizes the errors (in terms of KLD and RMSE) caused by loops.

The remainder of this paper is organized as follows. In Section 2, we describe cooperative localization using TRW-NBP. An empirical approach for finding ρ is presented in Section 3. Finally, Section 4 provides some conclusions and suggestions for future work.

2. COOPERATIVE LOCALIZATION BASED ON TREE-REWEIGHTED NONPARAMETRIC BELIEF PROPAGATION

2.1. The Localization Problem

Consider N_a anchors and N_t targets scattered randomly in a planar region, and denote the two-dimensional (2D) location of node t by x_t . The target u obtains a noisy measurement d_{tu} of its distance from node t with some probability $P_d(x_t, x_u)$:

$$d_{tu} = \|x_t - x_u\| + v_{tu}, \quad v_{tu} \sim p_v. \quad (1)$$

For simplicity, we assume the noise v_{tu} has a zero-mean Gaussian distribution, and ideal model for probability of detection:

$$P_d(x_t, x_u) = \begin{cases} 1, & \text{for } \|x_t - x_u\| \leq R, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

where R represents transmission radius. We will indicate with the binary variable o_{tu} whether an observation is available ($o_{tu} = 1$) or not ($o_{tu} = 0$). Finally, each node t has some prior distribution denoted $p_t(x_t)$. The joint distribution is given by:

$$p(x_1, \dots, x_{N_t}, \{o_{tu}\}, \{d_{tu}\}) = \prod_{(t,u)} p(o_{tu} | x_t, x_u) \prod_{(t,u)} p(d_{tu} | x_t, x_u) \prod_t p_t(x_t). \quad (3)$$

Vladimir Savic is supported by the FPU fellowship from Spanish Ministry of Science and Innovation. We also thank partial support by: ICT project FP7-ICT-2009-4-248894-WHERE-2, the Spanish National Project under Grant TEC2009-14219-C03-01, program CONSOLIDER-INGENIO 2010 under grant CSD2008-00010 COMONSENS, and National Project M3HF under Grant TEC2007-67520-C02-01/02/TCM.

Our objective is to compute (or approximate) the marginal beliefs $p(x_t | \{o_{tu}\}, \{d_{tu}\})$, for every target t . Then, we can easily estimate the positions, e.g., as mean values of these marginals.

2.2. Graph Representation

The relationship between a graphical model and a distribution $q(x_1, \dots, x_{N_t})$ may be represented in terms of potential functions ψ which are defined over graph's cliques. A clique (C) is a subset of nodes such that for every two nodes in C , there exists a link connecting the two. So the joint distribution is given by:

$$q(x_1, \dots, x_{N_t}) \propto \prod_{\text{cliques } C} \psi_C(\{x_i : i \in C\}). \quad (4)$$

We can now define potential functions which can express the joint posterior distribution (3). This only requires potential functions defined over variables associated with single nodes and pairs of nodes. Single-node potential represents the prior information, i.e., $\psi_t(x_t) = p_t(x_t)$. The pairwise potential between nodes t and u , which is defined as likelihood function $\psi_{tu}(x_t, x_u) = p(o_{tu} | x_t, x_u) p(d_{tu} | x_t, x_u)$, is given by:

$$\psi_{tu}(x_t, x_u) = \begin{cases} P_d(x_t, x_u) p_v(d_{tu} - \|x_t - x_u\|) & \text{when } o_{tu} = 1 \\ 1 - P_d(x_t, x_u) & \text{when } o_{tu} = 0 \end{cases} \quad (5)$$

Finally, the joint posterior distribution is given by:

$$p(x_1, \dots, x_{N_t} | \{o_{tu}, d_{tu}\}) \propto \prod_t \psi_t(x_t) \prod_{t,u} \psi_{tu}(x_t, x_u). \quad (6)$$

By marginalizing this joint distribution, we can find the true belief of each node. Exact marginalization is intractable, which is why we resort to near-optimal message-passing methods.

2.3. Tree-Reweighted Belief Propagation (TRW-BP)

In the standard TRW-BP algorithm the belief at a node t is proportional to the product of the local evidence at that node $\psi_t(x_t)$, and all *reweighted* messages coming into node t :

$$M_t(x_t) \propto \psi_t(x_t) \prod_{u \in G_t} m_{ut}(x_t)^{\rho_{ut}}, \quad (7)$$

where x_t is a state of node t , $\rho_{tu} = \rho_{ut}$ is the appearance probability of the edge (t, u) , and G_t denotes the neighbors of node t . The messages are determined by the message update rule:

$$m_{ut}(x_t) \propto \int \psi_u(x_u) \psi_{tu}(x_t, x_u)^{1/\rho_{tu}} \prod_{k \in G_u \setminus t} \frac{m_{ku}(x_u)^{\rho_{ku}}}{m_{tu}(x_u)^{1-\rho_{tu}}} dx_u, \quad (8)$$

where $\psi_{tu}(x_t, x_u)$ is the pairwise potential between nodes t and u . On the right-hand side, there is a product over all reweighted messages going into node u except for the one coming from node t . The update-rule (8) is carried out over the network. Upon convergence, the beliefs are computed through (7). In practice, it is more convenient to compute the beliefs at every iteration i . This leads to an equivalent form of TRW-BP: by replacing (7) in (8), we find that the belief equations and message-update rule of TRW-BP are, respectively, given by:

$$M_t^{(i)}(x_t) \propto \psi_t(x_t) \prod_{u \in G_t} m_{ut}^{(i)}(x_t)^{\rho_{ut}} \quad (9)$$

$$m_{ut}^{(i)}(x_t) \propto \int \psi_{ut}(x_t, x_u)^{1/\rho_{ut}} \frac{M_u^{(i-1)}(x_u)}{m_{tu}^{(i-1)}(x_u)} dx_u. \quad (10)$$

We can now apply TRW-BP to the localization problem. In the first iteration of this algorithm it is necessary to initialize $m_{ut}^1 = 1$ and $M_t^1 = p_t$ (i.e., information from anchors, if any) for all u, t , and then repeat computation using (9) and (10) until convergence or a preset number of iterations is attained. In a practical implementation, we have to use nonparametric version of TRW-BP (TRW-NBP). Hence, the beliefs and message update equations, (9) and (10), are performed using particle-based approximations [3].

2.4. Edge Appearance Probabilities

We will now describe how valid values for ρ_{tu} can be found. Given a graph G , let S be the set of all spanning trees T over G . Let $\vec{\rho}$ be a distribution over all spanning trees, i.e., a vector of non-negative numbers such that

$$\vec{\rho} \triangleq \{\rho(T), T \in S \mid \rho(T) \geq 0, \sum_{T \in S} \rho(T) = 1\}. \quad (11)$$

Observe that there are many such distributions. For a given $\vec{\rho}$ and a given (undirected) edge (t, u) , $\rho_{tu} = P_{\vec{\rho}}\{(t, u) \in T\}$, i.e., ρ_{tu} is the probability that the edge (t, u) appears in a spanning tree T chosen randomly under $\vec{\rho}$. Thus, ρ_{tu} represents *edge appearance probability* of the edge (t, u) . A valid collection of edge appearance probabilities must correspond to a valid distribution over spanning trees. For instance, $\rho_{tu} = 1$ for all edges, is not a valid collection of edge appearance probabilities, unless the graph G is a tree.

Finding a valid collection ρ_{tu} is difficult since there is a large number of spanning trees even in small graphs. For example, in 4-node clique there are 16 different spanning trees (Figure 1), and each edge appears exactly in 8 of them. Observe that if $\vec{\rho}$ is uniform over all spanning trees, then $\rho_{tu} = 0.5$ for every edge. Discovering all spanning trees, choosing a good $\vec{\rho}$, and then computing all ρ_{tu} would require significant network overhead, even for small networks. In [7], an alternative option is described, based on searching for trees (not necessarily spanning trees) in G . In any case, determining a valid collection ρ_{tu} requires a procedure similar to routing, which we prefer to avoid for our distributed inference problem.

We note that the choice $\rho_{tu} = 1$ for all edges corresponds to standard BP. In TRW-BP on graphs with cycles, it is easy to see that $\rho_{tu} \leq 1$ for all edges. Hence, by removing the restriction of valid ρ_{tu} and making ρ_{tu} uniform, we intuit that we can combine the benefits of BP (distributed implementation) and of TRW-BP (improved performance).

3. OPTIMIZING THE EDGE APPEARANCE PROBABILITY

From now on, we will consider uniformly reweighted NBP, i.e., $\rho_{tu} = \rho$ for all edges. We will now evaluate the impact of ρ on TRW-NBP through Monte Carlo simulation. We will first consider a small-scale network with 4 nodes, for which we can compute the true marginal posteriors. From this network, we will draw some important conclusions necessary for larger networks. Then, we determine the optimal ρ , with respect to transmission radius, in grid and random topologies. Due to the high computational cost of learning the optimal ρ in a 2D space, we will focus on 1D localization in the simulations. We use the following parameters: standard deviation of Gaussian noise is 30 cm, 200 particles per message, and 8 iterations. Finally, all results represent the average over 200 Monte Carlo runs.

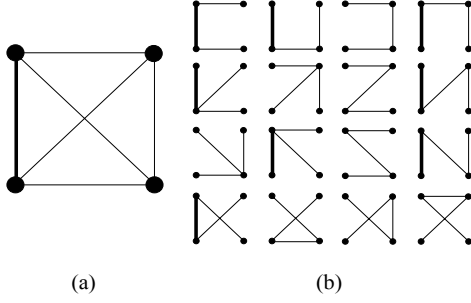


Fig. 1: (a) 4-node clique, and (b) 16 spanning trees. Each edge (e.g., bolded edge) appears exactly in 8 out of 16 spanning trees, so $\rho = 0.5$ for each edge, under a uniform distribution over the spanning trees.

3.1. A 4-Node Clique

We consider fully-connected network with 4 targets in 1D space (see Figure 1a for 2D case). In addition, there are 4 anchor nodes (not depicted), each of them connected exactly to one target. Our goal is to estimate the true belief, TRW-NBP beliefs and estimated locations. The latter are given by the minimum mean square error (MMSE) estimate from the belief. We run TRW-NBP for different values of ρ and, for each result, we compute KLD between true and TRW-NBP beliefs, and RMSE of estimated locations, all shown in Figure 2. According to Figure 2, we can make the following conclusions:

1. Both RMSE and KLD reach the minimum for the same $\rho < 1$. That means that it is sufficient to use only RMSE for learning the optimal ρ in larger networks, where the computation of true beliefs (necessary for computing KLD) is intractable.
2. The optimal ρ (ρ_{opt}) is 0.5, which is the same as the theoretical value (Figure 1b), under a uniform distribution over spanning trees. NBP ($\rho = 1$) performs worse than optimum TRW-NBP in terms of both KLD and RMSE. For a comparison between the three different beliefs, see Figure 3.
3. A wide range of ρ (in our example, 0.4-1) provides better performance than NBP in terms of both KLD and RMSE. That means that we can even use a coarse approximation of ρ_{opt} .
4. The RMSE is rather insensitive to ρ , for $\rho > \rho_{\text{opt}}$. Hence, care needs to be taken when interpreting RMSE figures as a function of ρ , as the effect on KLD may be much more pronounced.¹

Taking these conclusions into account, we now move on larger networks.

3.2. Grid and Random Topology Networks

We consider a network with 25 target nodes and 4 anchors in a 20m wide deployment area. We consider different values of the communication range² R , and the edge appearance probability ρ .

For the grid topology (where the distance between neighboring nodes is 0.6 m), Figure 4a shows the RMSE as a function of ρ , with

¹This could be a problem for learning in larger networks, where it is practically impossible to obtain smooth curves. However, we always use only confident digits (by rounding-up RMSE) and, in case of more minimums, we use the minimum corresponding to the lowest value of ρ (e.g., see Figure 5a). This approximation still keeps KLD quite close to minimum.

²The values of R are chosen so as to provide the same average node degree (n_d) both for grid and random topology.

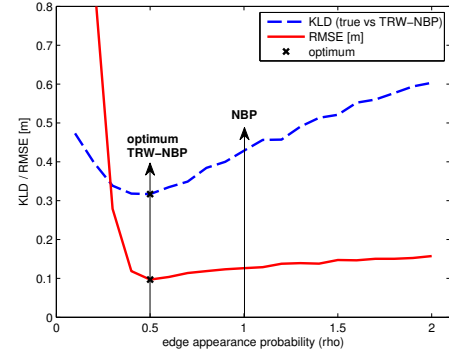


Fig. 2: Optimum ρ estimation in 4-node network.

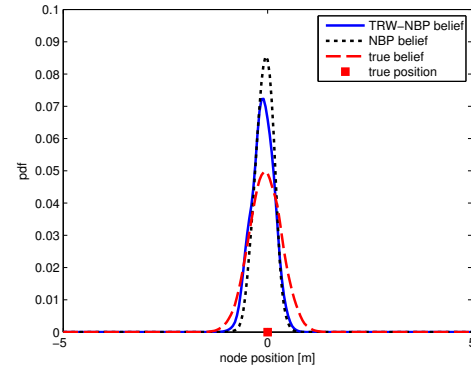


Fig. 3: NBP, TRW-NBP ($\rho = 0.5$), and true belief for the one of the target nodes ($x = 0$ m).

parameter n_d . We mark the optimal ρ , for each distinct value of n_d . This allows us to plot ρ_{opt} as a function of n_d (see Figure 4b). We observe that ρ_{opt} decreases nearly exponentially with n_d . Hence, we fit $\rho_{\text{opt}}(n_d)$ as

$$\rho_{\text{opt}}(n_d) = \rho_0 \cdot e^{-k_p n_d}, \quad (12)$$

where parameters $\rho_0 = 3.187$ and $k_p = 0.199$ are found using least-square fitting. We did the same test for random topology (Figure 5), and obtained: $\rho_0 = 2.656$ and $k_p = 0.161$. Note that for random topology, it is harder to obtain sufficient statistics (Figure 5a), so the fitting is less confident compared with the grid topology. In any case, we conclude the following:

1. The difference between coefficients for random and grid topology is small, which means that the value of ρ_{opt} depends more on the average node degree than the particular network configuration.
2. Though tempting to state that choosing $\rho = 1$ will lead to similar performance as $\rho = \rho_{\text{opt}}$, due to the almost flat curves for $\rho > \rho_{\text{opt}}$, this statement is not true when the performance is measured in terms of KLD (see Figure 2).

As an aside, when n_d becomes very small, the fitted value for ρ_{opt} can be larger than 1. This is merely a side-effect of the fitting. In practice, when $\rho_{\text{opt}} > 1$, one should set $\rho_{\text{opt}} = 1$.

Finally, it is worth noting that n_d can easily and quickly be found using a consensus algorithm [8]. In that case, the computational/communication cost will be nearly the same as for NBP.

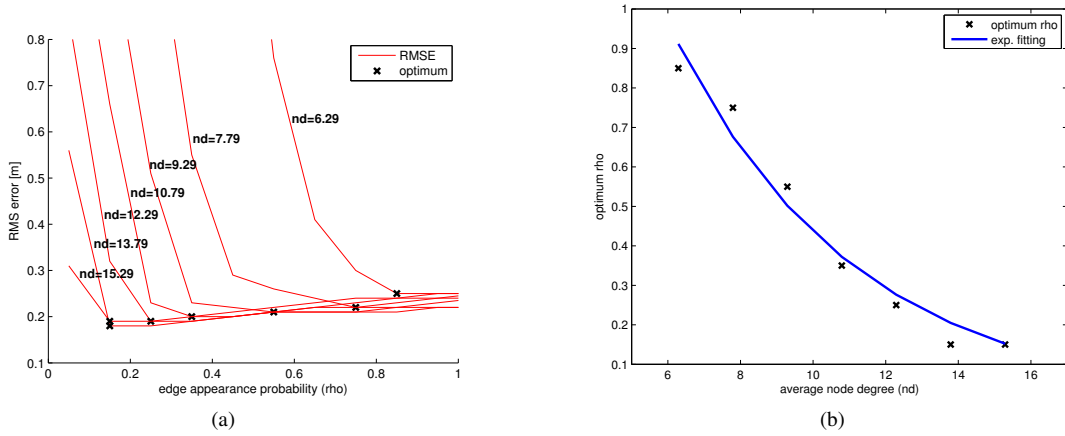


Fig. 4: Grid topology: (a) RMSE for different transmission radius, (b) Empirical model for optimum ρ .

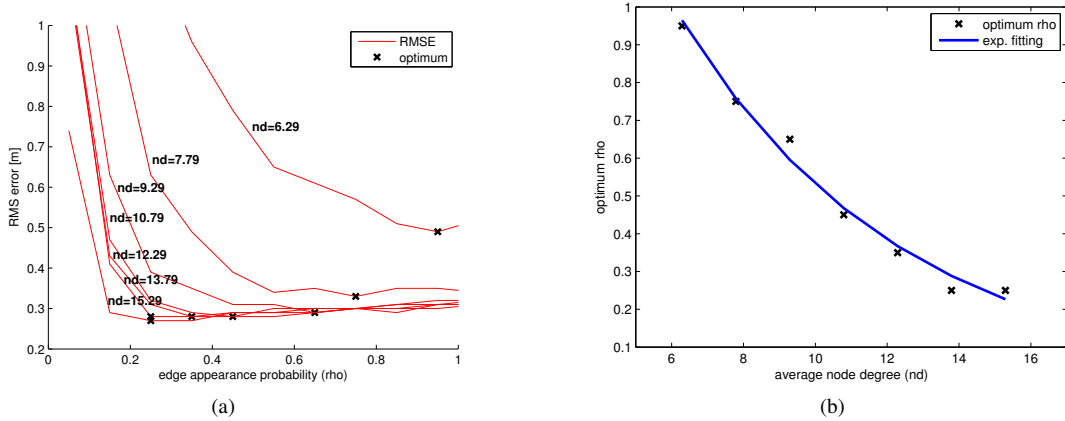


Fig. 5: Random topology: (a) RMSE for different transmission radius, (b) Empirical model for optimum ρ .

4. CONCLUSIONS

We presented a cooperative localization algorithm based on tree-reweighted nonparametric belief propagation (TRW-NBP), which combines the distributed nature of NBP and the improved performance of TRW-BP. In contrast to TRW-BP, we propose to use a constant edge appearance probability ρ . In contrast to NBP, we propose to set $\rho < 1$. Through Monte Carlo simulations, we have verified performance gains in terms of RMSE and KLD w.r.t. the true distribution. We have found that (i) the optimal ρ_{opt} does not depend on the particular criterion (RMSE or KLD); (ii) ρ_{opt} decreased in networks with more loops; (iii) ρ_{opt} can be expressed as a simple function of the average node degree (n_d). Our future work will focus on the extension to 2D and 3D space. Preliminary results (not reported in this paper) indicate similar behavior of ρ_{opt} as a function of n_d . Our final goal is the implementation and evaluation of different variations of belief propagation on wireless motes.

5. REFERENCES

- [1] A.T. Ihler, J. W. Fisher III, R. L. Moses, and A. S. Willsky, "Nonparametric belief propagation for self-localization of sensor networks", *IEEE Journal On Selected Areas In Communications*, vol. 23, issue 4, pp. 809-819, April 2005.
- [2] J.S. Yedidia, W.T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations", *Exploring artificial intelligence in the new millennium*, vol. 8, pp. 239-269, Jan. 2003.
- [3] V. Savic, A. Poblacion, S. Zazo, and M. Garcia, "Indoor Positioning using Nonparametric Belief Propagation based on Spanning Trees", *EURASIP Journal on Wireless Communications and Networking*, 2010.
- [4] M.J. Wainwright, T.S. Jaakkola and A.S. Willsky, "Tree-reweighted belief propagation algorithms and approximate ML estimation by pseudo-moment matching", in *Workshop on Artificial Intelligence and Statistics*, Jan. 2003.
- [5] H. Wymeersch, J. Lien and M.Z. Win, "Cooperative localization in wireless networks", in *Proc. of IEEE*, vol. 97, issue 2, pp. 427-450, Feb. 2009.
- [6] Z.I. Botev, "A Novel Nonparametric Density Estimator", Technical Report, The University of Queensland, Australia, 2006.
- [7] V. Kolmogorov "Convergent Tree-reweighted Message Passing for Energy Minimization", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1568-1583, Oct. 2006.
- [8] R. Olfati-Saber, R.M. Murray, "Consensus problems in networks of agents with switching topology and time-delays", in *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520-1533, Sept. 2004.